DETECTION OF INSECTS IN BULK WHEAT SAMPLES WITH MACHINE VISION

I. Y. Zayas, P. W. Flinn

ABSTRACT. *Digital imaging techniques were used to identify insects and body parts (adult lesser grain borer) of Rhyzopertha dominica beetles in bulk wheat samples. The samples contained dockage (weed seeds and damaged wheat kernels), which made the identification of insects more difficult. Multispectral analysis (Red, Green, Blue) was used in combination with pattern recognition techniques. Insect identification experiments were done with subimages of 8 × 8 pixels. Statistical multivariate analysis and structural patterns were used for finding insects. The position of the insect (ventral, dorsal, side) and particles clinging to the insect affected recognition success. Recognition success of R. Dominica adults, some grass seeds, and other non-wheat components was higher than 90%.*

Keywords. *Machine vision, Insects, Wheat, Pattern recognition, Multivariate analysis, Digital imaging.*

Infestation of grain is a serious problem in marketing channels, causing big losses and health hazards. A variety of factors affect wheat value. One of these factors is the presence of live insects. Manis (1992) described the sampling, inspecting and grading process.

According to current U.S. grain standards, the special grade designation of "infested" is given when "The representative sample (other than shiplots) contains two or more live weevils, or one live weevil and one or more other live insects injurious to stored grain, or two or more live insects injurious to stored grain", (*Official U.S. Standards for Grain*, 7 CFER part 810.107, 1997).

Many different methods are used to sample wheat: diverter-type mechanical samplers, Ellis cup, pelican, and the compartmented probe (trier). After the grain sample is obtained, it is taken to the laboratory and divided into a work sample, a file sample and a moisture sample. The work sample is used by the inspector to determine its grade. The first factors the inspector checks for are odor, infestation, and other unusual conditions. The whole work sample is checked visually for live insects. Insects are separated from the grain with official grain sieves. Insects and other fine material are collected in a bottom pan when the grain is shaken over the sieve. Insects are counted and identified in the grain pan. This can be a laborious procedure, and often insects are missed due to human error.

A possible method that may be used to detect insects in grain is acoustic detection (insect movement and feeding sounds are detected with highly amplified microphones in grain), reported by (Hagstrum, 1993). Among other alternative methods reported are nuclear magnetic resonance spectroscopy (NMR) (Chambers et al., 1984), near infrared spectroscopy (NIR) (Ridgway and Chambers, 1996) and X-radiation (Schatzki and Keagy, 1991; Keagy and Schatzki, 1993).

Machine vision is a rapidly developing technique. Pattern recognition methodology in combination with machine vision is utilized in many areas of industry and agriculture for control and inspection purposes. Some research has been done on the use of imaging techniques for descriptive morphometry of individual insects, such as termites (Grace et al., 1986), bees (Batra, 1988) and mosquitoes (Zhou et al., 1985), but no studies on bulk samples have been reported. Imaging techniques, using photomultipliers and pattern recognition techniques were used for recognition of several insects, prevailing in a cotton field (Atmar et al., 1973). The study looked at individual objects in the field of view and obtained a recognition rate of approximately 85%.

There is no published research on detection of insects in the bulk grain, using digital imaging technique. Machine vision with pattern recognition technique is an attractive approach for grain quality inspection. It is an objective method, it may exclude the tedious and laborious process of grain inspection for presence of insects. Successfully developed algorithm may give insects count, a picture of detected insects for visual confirmation and archiving of samples if needed.

OBJECTIVES

The objective of the study was to explore a possibility of developing a method to detect whole insects or body parts in the bulk samples of grain and dockage, using machine vision. The main goal of the study was to determine x,y coordinates of a subimage, belonging to insect versus non-insect elements of the image. The search for x,y coordinates was based on multispectral measurements, extracted from Red, Green and Blue images for the same sample. Subimages were used instead of individual pixels, to reduce computational time. A goal for image processing

Article was submitted for publication in September 1997; reviewed and approved for publication by the Information & Electrical Technologies Div. of ASAE in March 1998. Presented as ASAE Paper No. 97-3149.

Mention of a trade name, proprietary product, or specific equipment does not constitute a guarantee or warranty by the USDA and does not imply approval of a product to the exclusion of others that may be suitable.

The authors are **Inna Y. Zayas, *ASAE Member Engineer***, Electronics Engineer, and **Paul W. Flinn,** Research Biologist, USDA-ARS, GMPRC, Manhattan, Kans. **Corresponding author:** I. Y. Zayas, USDA-ARS, GMPRC, 1515 College Av., Manhattan, KS 66502; tel: (913) 776-2758; fax: (913) 776-2792; e-mail: iz@usgmrl.ksu.edu.

algorithm development was to determine, what subimage size should be used to efficiently run a program, to reduce amount of information to process. Accuracy of determining insect body boundaries was another challenging goal, considering different angles of insects or insect parts position in a sample and subimage size. The similarity in color of insects to some components of the dockage made the task more difficult. An objective was to develop effectively working algorithm despite clinging particles of dust, which obscure insect tissue and change multispectral response.

MATERIAL

Grain samples used in this study came from a field study reported by Flinn et al. (1996). They were sievings from hard red winter wheat that had become infested with several species of stored product insects. Bins were infested with *Rhyzopertha dominica* (lesser grain borer), *Cryptolestes ferrugineus* (rusty grain beetle), *Cephalonomia waterstoni* (parasitoid wasp), and *Choetospila elegans* (parasitoid wasp). However, other species, *Oryzephilus surinamensis* (sawtoothed grain beetle) and *Triboleum castaneum* (red flour beetle) were also found in the grain. The predominant species was the lesser grain borer, (*Rhyzopertha dominica*).

A pneumatic grain sampler (Probe-A-Vac, Cargill, Minneapolis, Minn.) was used to obtain grain samples from a 35.2 m^3 bin of hard red winter wheat. Samples were obtained using a vacuum probe from different locations in a bin. Samples were a blend of wheat kernels and dockage, including some grass seeds. The 3 kg samples were passed over an inclined sieve (89 × 43 cm, 1.6 mm aperture) to separate whole grain from dockage and insects. Some wheat kernels were added back to sieved material before image acquisition to determine, if wheat could be distinguished from insect parts.

IMAGE ACQUISITION

EQUIPMENT

Kontron (Kontron Electronic Gmbh) hardware and software were used for image acquisition. IBAS 2.5 software was installed on a 486 DX 33 MHz DOS computer, with 16 Mb of memory. Data processing and data visualization were done using a 75 MHz SPARC LX workstation. Images were obtained with a Progress 3000 color camera mount on a Bencher photostand.

A Progress 3000 color camera was calibrated to 1488 × 1180 pixels, one of the six possible calibrations. The chosen calibration was used to acquire images in a 512 × 512 pixel format, with a pixel aspect ratio of 1.06. The field of view was 5.4 mm × 5.3 mm. Calibration values were scale x: 1.05×10^{-2} mm = 1 pixel; y: 1.03×10^{-2} mm = 1 pixel. Samples were placed in a crate cell, as shown in figure 1. The depth of the sample was approximately 5 mm and the bottom of the crate cell was completely covered. Color calibration of Progress camera is a part of calibration procedure, a program supplied by manufacturing company Kontron. Color calibration procedure is programmed for sampling white and black pixels. Samples were prepared in a manner, that one or more insects should be present in the field of view. Samples were shaken to expose insects. Some dockage components, such as grass seeds, were difficult to differentiate from insects, because of similarity of color and shininess. The position of the insect (ventral, dorsal, side) and particles clinging to the insect affected recognition success. The best quality image, considering the problems mentioned above, was achieved with backlight illumination, one halogen lamp 250 W. To reduce reflection, a dome with flat matte white paint inside was placed over the sample. The Progress 3000 camera was placed above the dome with a Nikkor 50 mm lens inside the dome. Images of seventy samples were stored for future analysis.

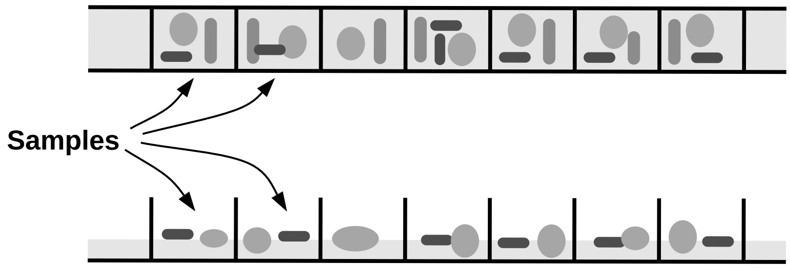
DATA ANALYSIS AND RESULTS

A block diagram, illustrating the data processing sequence, is shown in figure 2. Red, Green, Blue, hue, saturation and lightness features were extracted from histograms for each of the subimages (8 × 8 pixels). Thirty features from subimage histograms included the mean, standard deviation, skewness, kurtosis, and several derived features. Five derived features were: (red mean/green mean), (red mean/blue mean), (blue standard deviation/ green standard deviation), (red standard deviation/green standard deviation), and squared (green standard deviation/red standard deviation) (table 1).

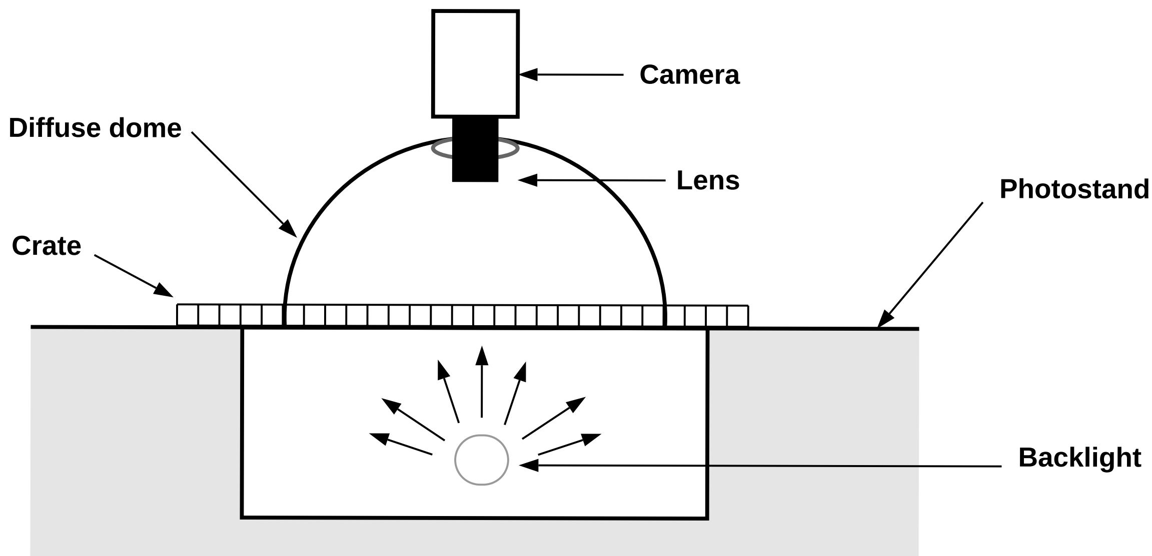
Multivariate analysis was used to classify observations, which were subimages of insects, grass seeds and others for two and three class categorization. Discriminant analysis was run using a program written inhouse, Zayas et al., (1996) and SAS procedure (SAS, 1991) for verification of discriminant analysis results. In the training and testing steps, the program computed the probability of each subimage belonging to a certain class. The results were presented as percent of correctly identified subimages, with x,y coordinates of subimage location stored. For two-class analysis, categorized observations were insect versus non-insect subimages. For three-class, categorized observations were subimages from insects, grass seeds, and others elements of an image. Grass seeds were chosen from other components of the dockage, because they were very close in color to insects. “Others” included all kinds of dockage and shaded areas in between objects in the field of view.

Observations for training and testing were selected in two different ways. In the first approach, training and testing data sets were created by interactive selection of representative subimages from three classes: insects (3932), grass seeds (2432), and others (2244). The pool of subimages was randomly divided in half for training and test data sets for each class. Discriminant and canonical analysis procedures of the SAS statistical package (SAS, 1991) and an inhouse written program were run to compute linear and quadratic functions with different features models. Discriminant analysis was also run in conjunction with canonical analysis; canonical functions were computed first, and the results were used as an input for the calculation of discriminant functions.

The STEPDISC procedure of SAS was used to choose the number of features in the model and the best performing features. This procedure ranks the features by R^2. The output of the procedure shows the value of the Average Squared Canonical Correlation (ASCC). Lack of



(a) Sample crate.



(b) Lighting chamber.

Figure 1–Block diagram of data processing.

significant changes in the values of the ASCC shows, that adding features to the model does not improve performance of the model (table 2). Comparison of the performance of different features models was done using Linear Discriminant Analysis for two-class and three-class categorization. The results are given in the bar plots in figure 3. Table 2 shows that no drastic changes occur in the ASCC after the fifth feature, but figure 3 shows, that only starting from the 20 features model, recognition of insects and other classes reached a higher rate of approximately 92% for insects. The output of STEPDISC shows overall estimation for all classes and STEPDISC may be considered a helpful tool, but not a reliable tool for classification decision. The best performance for two-class categorization was achieved with the 35 features model, with a correct recognition rate of 91.3% for insects and 96.3% for others for an average of 93.8% (table 3 and fig. 3). The best performance for three-class categorization was achieved with the 25 features model, with a correct recognition rate of 91.3% for insects, 91.2% for others, and 94.7% for seeds for an average of 92.4%. For further testing, models with 20 features or more and higher were chosen.

Linear, Quadratic, and Canonical discriminant analysis methods were tested for better performance. Results of analysis for test data set for the three-class categorization with different numbers of features in a model, and different multivariate analysis methods are displayed in table 3. The best results were achieved when Canonical and Discriminant analysis was run with Quadratic functions computed with a 35 features model (average 92.6%, 90.3% insects, 93.1% others, and 94.6% grass seeds). Slightly worse results were achieved for test data, when Discriminant Linear Functions were computed with the 25 features model (average 92.4%, 91.3% insects, 91.2% others, and 94.7% grass seeds).

The best results for the test data set for two-class categorization were achieved when Canonical and Discriminant analysis with Quadratic functions were computed with the 35 features model (average 94.4%, 92.5% insects, and 96.3% others). Canonical plots were created to visualize the performance of canonical analysis. Figure 4 shows the results of categorization by two canonical functions for three-class (insects, dark grass seeds, and others) and two-class discrimination (insects and non-insects). Clustering of the observations into three groups and two groups with some overlap may be observed on these plots.

The second approach to verify algorithm performance used binarization of images (fig. 5). Binarization of the acquired subimages was used to verify visually, which parts of the image were identified correctly as a certain class. The

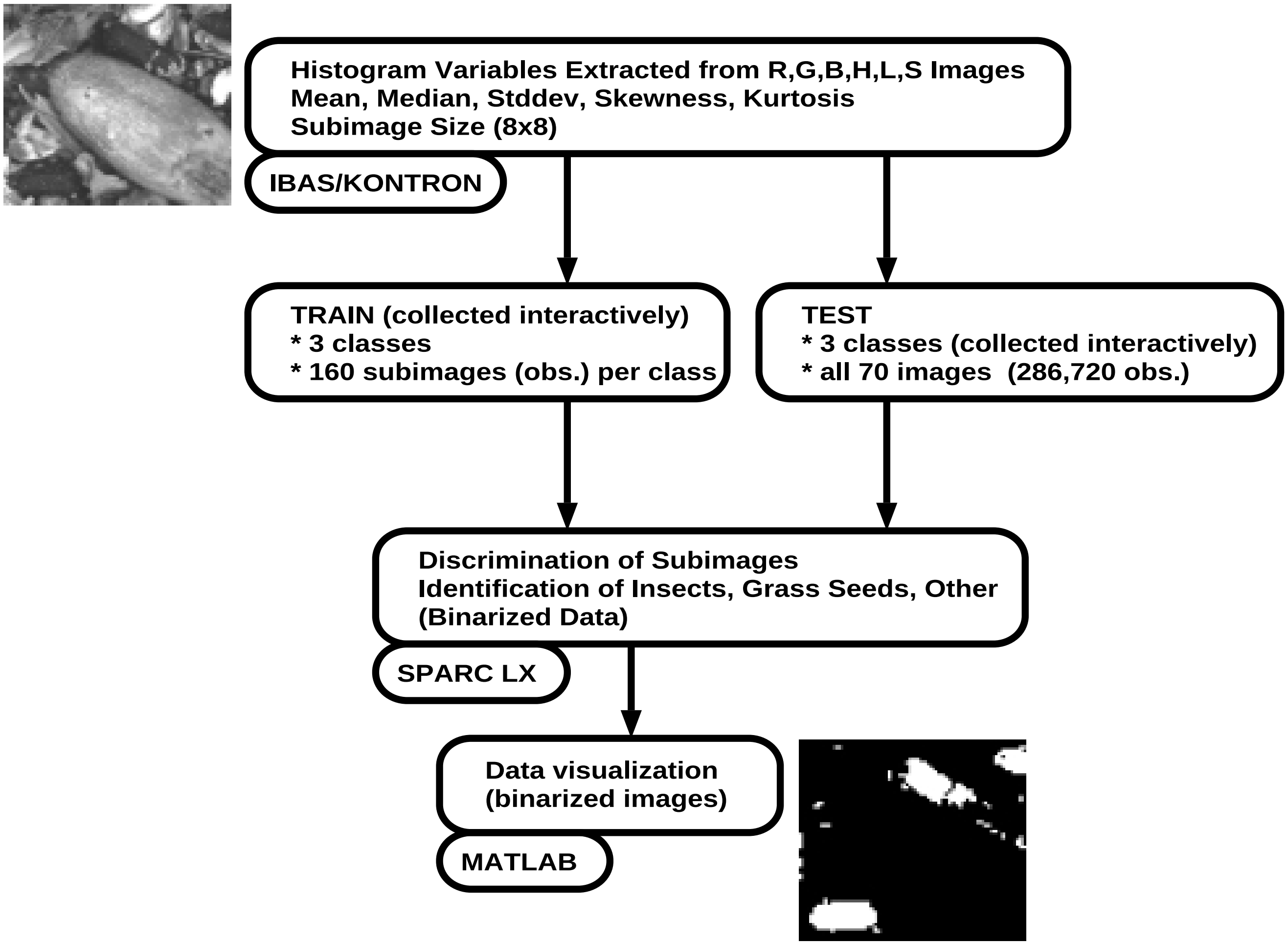


Figure 2–Sketch of experiments scene.

Table 1. Variables (35) used for identification of insect tissue (8 × 8 subimage)

Histogram Statistics of Subimage Red, Green, Blue, Hue, Lightness, Saturation		Derived Variable	Definition*
MEDIAN	- Median	A	RMEAN / GMEAN
MEAN	- Mean	B	RMEAN / BMEAN
STDDEV	- Standard Deviation	C	BSTDDEV / GSTDDEV
SKEWNESS	- Skewness	D	RSTDDEV / GSTDDEV
KURTOSIS	- Kurtosis	M_D_D	$(GSTDDEV / RSTDDEV)^2$

* R, G, B, are added to related histogram statistics terms.

output of discriminant functions was a data set with each subimage assigned to a certain class. Each subimage carried information about x,y coordinates of the subimage location. The recognition algorithm allowed comparison of a binary image with the original gray scale image. The program assigned subimage pixels belonging to insects to "1" (white), and the rest of the image to "0" (black), to produce binary images. MATLAB (MATLAB, 1993) software was used for data visualization to display images for verification. Subimages of 8 × 8 pixels were chosen for final analysis and binarization. The 35 features model performed reasonably well, although further refining of the algorithm was needed to reduce noise for better insect discrimination from the rest of the image. In some cases, it was difficult to verify binarized images with incomplete insect bodies. Insect body parts such as open wings, head or leg were difficult to verify in binary images. Differentiation of insects from wheat did not pose a problem.

CONCLUSION

The application of machine vision for finding insects in sieved grain samples was studied. Multispectral information from histograms of Red, Green and Blue, hue, saturation and lightness images, for 8 × 8 pixels subimages, was used as an input to multivariate statistical analysis to locate subimages belonging to insects or non-insects components of the samples. The developed algorithm for two-class discrimination of subimages: insect versus non-insect showed the best performance when Canonical and Discriminant analysis was run with Quadratic functions computed, using 35 features model. For three-class discrimination of subimages, the same algorithm and 35 features model produced satisfactory results for discriminating insect parts from the wheat and dockage. Grass seeds similar in color to insects were components of the dockage. Despite the fact that some elements of dockage such as dark grass seeds, were difficult to discriminate from insects, insects were discriminated reasonably well. It was difficult to discriminate separate insect body parts from other classes. Position of the insects (ventral, dorsal, side), and particles clinging to the insect, affected recognition success of the insects versus the rest of the sample components. Particles of dust clinging to the surface of the insects made recognition of insects more difficult. Some separate insect body parts were present in the image in various angular positions among dockage

Table 2. The SAS System stepwise discriminant analysis

Step	Variable*	Partial R^2	Average Squared Canonical Correlation
1	B	0.6425	0.3213
2	HMEAN	0.4927	0.5670
3	HSTDDEV	0.3568	0.6495
4	GMEDIAN	0.1421	0.6786
5	M_D_D	0.0921	0.6943
6	RMEDIAN	0.0458	0.7014
7	D	0.0517	0.7098
8	C	0.0494	0.7172
9	HSKEW	0.0294	0.7211
10	SMEAN	0.0250	0.7251
11	A	0.0278	0.7294
12	GSTDDEV	0.0146	0.7315
13	RSTDDEV	0.0225	0.7346
14	LMEAN	0.0164	0.7365
15	BMEAN	0.0644	0.7448
16	GMEAN	0.0943	0.7559
17	HMEDIAN	0.0120	0.7572
18	SSKEW	0.0125	0.7586
19	RMEAN	0.0119	0.7598
20	LSTDDEV	0.0095	0.7608
21	BSTDDEV	0.0096	0.7618
22	BMEDIAN	0.0106	0.7631
23	SKURT	0.0059	0.7639
24	SMEDIAN	0.0053	0.7646
25	LSKEW	0.0051	0.7651
26	BKURT	0.0169	0.7667
27	HKURT	0.0113	0.7678
28	RKURT	0.0103	0.7689
29	GSKEW	0.0034	0.7692
30	BSKEW	0.0017	0.7695
31	SSTDDEV	0.0022	0.7697
32	LKURT	0.0012	0.7698
33	GKURT	0.0028	0.7702
34	LMEDIAN	0.0011	0.7703
35	RSKEW	0.0006	0.7704

* See table 1; features taken from stepdisc output from February 1997 all results from SAS Discrim procedure.

Table 3. Results of discriminant analysis (test data)

Procedures†		Features‡	Three Class* Insects (%)	Three Class* Others (%)	Three Class* Grass Seeds (%)	Three Class* Avg	Two Class Insects (%)	Two Class Others (%)	Two Class Avg
Can & Disc	L	20	91.51	89.66	94.82	91.99	89.01	96.32	92.66
Can & Disc	Q	20	89.78	92.25	94.41	92.14	90.08	95.21	92.64
Can & Disc	L	25	91.00	91.44	95.07	92.50	88.71	96.83	92.77
Can & Disc	Q	25	90.28	92.60	94.41	92.43	90.59	95.51	93.05
Can & Disc	L	35	91.40	90.73	94.16	92.09	91.10	96.83	93.96
Can & Disc	Q	35	90.28	93.05	94.57	92.63	92.52	96.28	94.40
Discrim	L	20	91.76	89.84	94.90	92.16	88.86	96.49	92.67
Discrim	Q	20	94.10	79.59	96.88	90.19	96.08	83.92	90.00
Discrim	L	25	91.30	91.18	94.74	92.40	88.81	96.79	92.80
Discrim	Q	25	94.30	79.68	97.12	90.36	95.98	84.77	90.37
Discrim	L	35	92.12	90.37	94.33	92.27	91.25	96.32	93.78
Discrim	Q	35	93.79	82.53	96.71	91.01	96.19	86.74	91.46

* Number of observations per class: Three class—Insects, 1966; Others, 1122; Grass seeds, 1216; Two class—Insects, 1966; Others, 2338.

† Can = Canonical Analysis; Disc = Discriminant Analysis;Q = Quadratic; L = Linear.

‡ 20 features: rmedian, rmean, rstddev, gmedian, gmean, gstddev, bmean, hmedian, hmean, hstddev, hskew, lmean, lstddev, smean, sskew, a, b, c, d, m_d_d; 25 features: rmedian, rmean, rstddev, gmedian, gmean, gstddev, bmedian, bmean, bstddev, hmedian, hmean, hstddev, hskew, lmean, lstddev, lskew, smedian, smean, sskew, skurt, a, b, c, d, m_d_d.

particles, which made visual verification more difficult. The shiny surface of insects demanded special illumination settings to avoid blind spots in the image. The choice of subimages for the training data set affected results of recognition due to the above mentioned image problems (amount of dust, position angle, etc). Further image processing refinement of the algorithm is needed to reduce noise in images for identification of dead insects and their body parts.

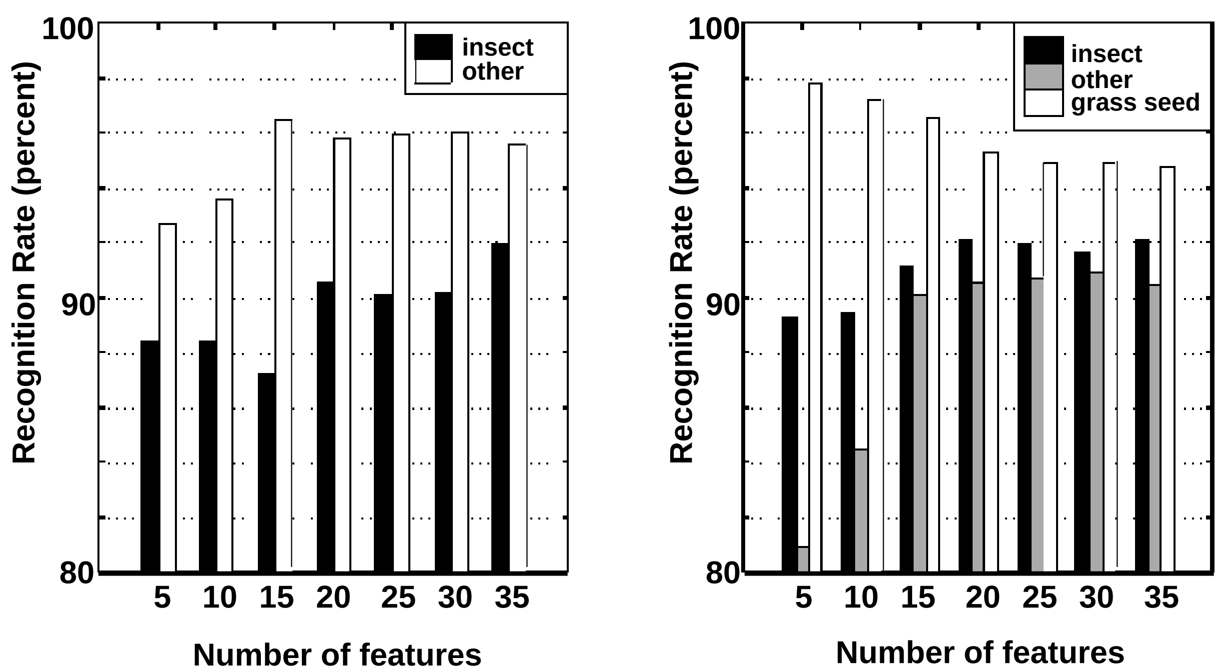


Figure 3–Bar plots of correct recognition rate for two and three classes for 5-35 features models, by Linear Discriminant Analysis.

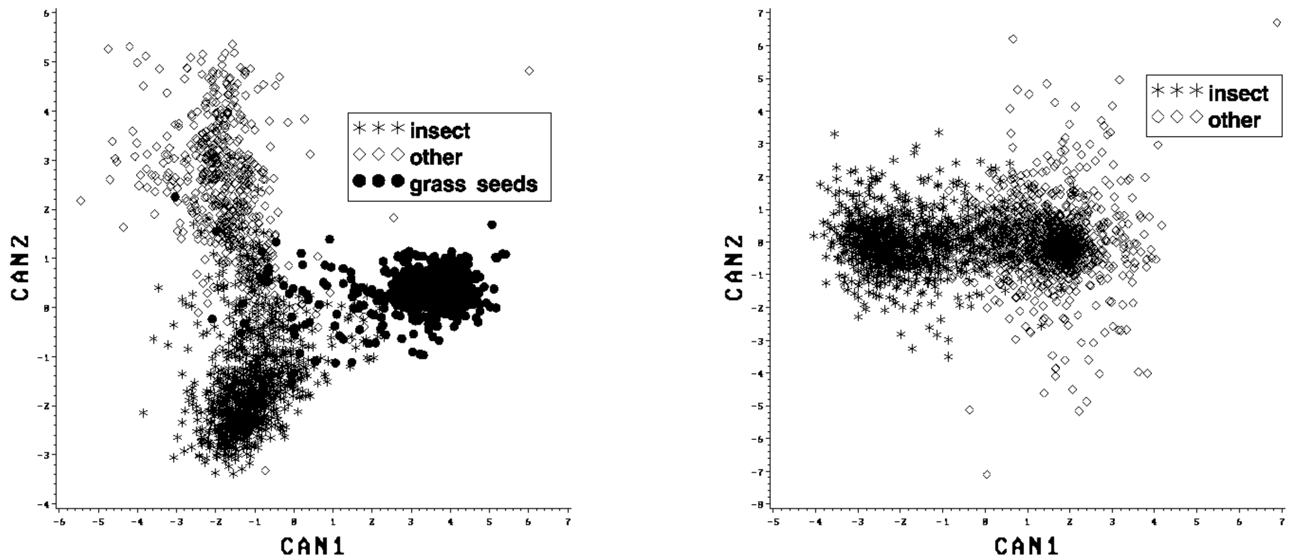


Figure 4–Canonical plot for discrimination of insects, grass seeds, others (8 × 8 pix), 35 features model, for three and two class categorization.

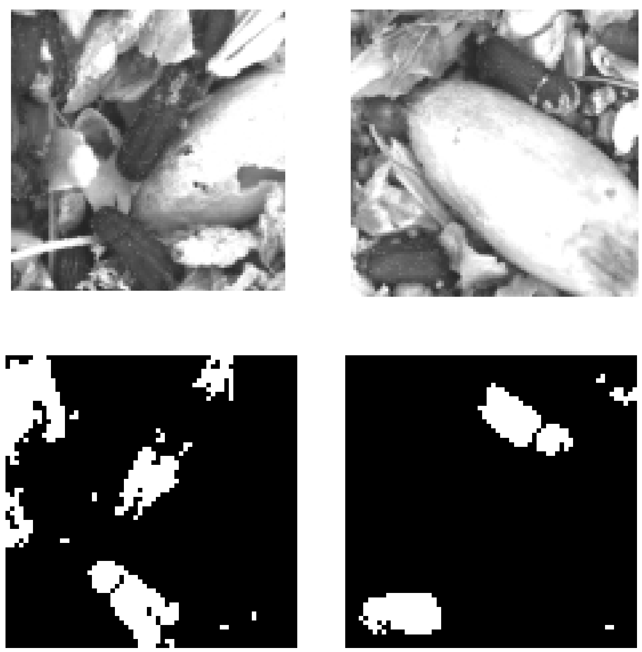


Figure 5–Finding insects by discriminant functions and image binarization.

ACKNOWLEDGMENT. The authors wish to thank A. Richter, a student of Kansas State University, for work with data acquisition and data analysis, and K. Friessen, a technician at Biological Research Unit, Grain Marketing Production and Research Center, for sample preparation.

REFERENCES

Atmar, J. W., J. L. Pooler, F. C. Webb, G. M. Flachs, and J. J. Ellington. 1973. Construction of a device to identify and count insects automatically. *Environ. Entomol.* 2(4):713-716.

Batra, S. W. T. 1988. Automatic image analysis for rapid identification of Africanized honey bees. In *Africanized Honey Bees and Bee Mites*, eds. Needham, G. H., R.E. Page Jr., M. Delfinado-Baker, and C. E. Bowman. Chichester, U.K.: Ellis Horwood Ltd.

Chambers, J., N. J. McKevitt, and M. R. Stubbs. 1984. Nuclear magnetic resonance spectroscopy for studying the development and detection of the grain weevil, *Sitophilus granarius* (L.) (*Coleoptera: Curculionidae*), within wheat kernels. *Bull. Entomol. Res.* 74:707-724.

Flinn, P. W., D. W. Hagstrum, and W. H. McGaughey. 1996. Suppression of beetles in stored wheat by augmentative releases of parasitic wasps. *Environ. Entomol.* 25(2):505-511.

Grace, J. K., I. Kubo, and S. E. Lindow. 1986. Measurement of termite (*Isoptera: Rhinotermitidae*) feeding on paper by video image analysis. *J. Entomol. Sci.* 21(1):10-15.

Hagstrum, D. W. 1993. Comparison of acoustical detection of several species of stored grain beetles (*Coleoptera: Curculionidae, Tenebrionidae, Bostrichidae, Cucujidae)* over a range of temperatures. *J. Econ. Entomol.* 86(4):1271-1278.

Keagy, P. M., and T. F. Schatzki. 1993. Machine recognition of weevil damage in wheat radiographs. *Cereal Chem.* 70(6):696-700.

Matlab Reference Guide. 1993. Natick, Mass.: The Math Works, Inc.

Manis, J. M. 1992. Sampling, inspecting, and grading, In *Storage of Cereal Grains and Their Products.*, ed. D. B Sauer. St. Paul, Minn.: AACC.

Ridgway, C., and J. Chambers. 1996. Detection of external and internal insect infestation in wheat by near-infrared reflectance spectroscopy. *J. Sci. Food Agric.* 71(2):251-264.

Schatzki, T. F., and P. M. Keagy. 1991. Effect of image size and contrast on the recognition of insects in radiograms. *SPIE Proc.* 1379:182-188.

Statistical Analysis Systems Institute Inc. 1991. Ver. 6.07 Ed. *SAS USERS Guide.* Gary, N.C.: SAS.

USDA-GIPS. 1997. *The Official United States Standards for Grain.* Grain Inspection, Packers and Stockyards Administration. Washington, D.C.: U.S. Dept. of Agriculture.

Zayas, I. Y., J. L. Steele, and A. Katcevich. 1996. Wheat classification using image analysis and crush-force parameters. *Transactions of the ASAE* 39(6):2199-2204.

Zhou, Y. H., L. B. Ling, and F. J. Rohlf. 1985. Automatic description of the venation of mosquito wings from digitized images. *Systematic Zool.* 34(3):346-358.